Myrsini Katsikatsou and Irini Moustaki

Pairwise likelihood ratio tests and model selection criteria for structural equation models with ordinal variables

Article (Accepted version) (Refereed)

Original citation:
Katsikatsou, Myrsini and Moustaki, Irini (2016) Pairwise likelihood ratio tests and model selection criteria for structural equation models with ordinal variables. Psychometrika . pp. 1-23. ISSN 0033-3123

DOI: 10.1007/s11336-016-9523-z

Pairwise likelihood ratio tests and model selection criteria for structural equation models with ordinal variables

Myrsini Katsikatsou*, Irini Moustaki

Abstract

Correlated multivariate ordinal data can be analysed with structural equation models. Parameter estimation has been tackled in the literature using limited-information methods including three-stage least squares and pseudo-likelihood estimation methods such as pairwise maximum likelihood estimation. In this paper, two likelihood ratio test statistics and their asymptotic distributions are derived for testing overall goodness-of-fit and nested models respectively under the estimation framework of pairwise maximum likelihood estimation. Simulation results show a satisfactory performance of type I error and power for the proposed test statistics and also suggest that the performance of the proposed test statistics is similar to that of the test statistics derived under the three-stage diagonally weighted and unweighted least squares. Furthermore, the corresponding, under the pairwise framework, model selection criteria, AIC and BIC, show satisfactory results in selecting the right model in our simulation examples. The derivation of the likelihood ratio test statistics and model selection criteria under the pairwise framework together with pairwise estimation provide a flexible framework for fitting and testing structural equation models for ordinal as well as for other types of data. The test statistics derived and the model selection criteria are used on data on 'trust in the police' selected from the 2010 European Social Survey. The proposed test statistics and the model selection criteria have been implemented in the R package `lavaan`[1].

Keywords: latent variable modelling; composite likelihood; underlying variable approach.

*The project has been supported by ESRC, grant ES/L009838/1.

[1]**Acknowledgements**: We thank Professor Yves Rosseel, the developer of the R package `lavaan`, for adopting our R code related to PML methodology and incorporating into `lavaan`.

1 Introduction

Ordinal scales are widely used in social sciences for measuring attitudes and behaviour. A variable with an ordered categorical scale is called an ordinal variable (Agresti, 2010). There are two main approaches for modelling categorical (binary and ordinal) observed variables with latent variables, namely the full information maximum likelihood approach (FIML) used in item response theory (e.g. Skrondal & Rabe-Hesketh, 2004; Bartholomew et al., 2011) and the limited-information approach used in structural equation modelling (SEM) (e.g. Jöreskog, 1990, 1994; Muthén, 1984). The latter uses first and second order statistics included in the univariate and bivariate likelihood functions. The limited information approach is adopted here. The general framework of structural equation modelling includes models for continuous variables, categorical variables, and mixtures of variables (Arminger & Küsters, 1988; Muthén, 1984), confirmatory factor analysis (Jöreskog, 1969), mixed effects analysis (Fan & Hancock, 2012), multi-group analysis (Jöreskog, 1971; Muthén, 1989), latent growth curve analysis (Bollen & Curran, 2006), and non-linear models (Jöreskog & Yang, 1996; Wall & Amemiya, 2000) as special cases. Estimation and testing remain important research topics when models involve non-normally distributed observed variables such as ordinal variables. Taking into account the ordinal nature of a variable can result in a more accurate and powerful analysis as is pointed out by Agresti (2010). Jöreskog (2002) also recommends that ordinal variables should be analysed as such since they do not have origins or measurement units and consequently, means, variances, and covariances of ordinal variables do not have meaning.

In SEM, each observed ordinal variable is generated by an underlying continuous variable assumed to be normally distributed. Thus, FIML estimation requires the evaluation of normal probabilities of dimension equal to the number of the observed ordinal variables (Lee et al., 1990a; Poon & Lee, 1987). This renders FIML computationally infeasible when the number of ordinal variables is large. As a result, two- and three-stage limited-information least squares (3S-LS) estimation and testing theory have been proposed in the literature (Jöreskog, 1990, 1994; Jöreskog & Sörbom, 1996; Lee et al., 1990b, 1992; Muthén, 1984; Satorra, 2000; Satorra & Bentler, 2010, 1988; Asparouhov & Muthén, 2006, 2010) and implemented in software such as LISREL (Jöreskog & Sörbom, 1996), Mplus (Muthén & Muthén, 2010), EQS (Bentler, 2006), and the R package `lavaan` (Rosseel, 2012; Rosseel et al., 2012). Bayesian estimation methods of estimation, testing and model selection have also been developed (see e.g. Ansari & Jedidi, 2000, 2002; Lee, 2007; Palomo et al., 2007; Raftery, 1993, and references therein).

A competitive limited information estimation method is the pairwise maximum likelihood (PML) (Jöreskog & Moustaki 2001; De Leon 2005; Liu 2007; Katsikatsou et al. 2012; Katsikatsou 2013; Xi 2011). PML, similarly to 3S-LS, utilizes information from lower order margins (bivariate). It is a limited information estimation method that has been developed within the maximum likelihood (ML) estimation framework. Although PML estimation has been well developed in the literature of SEM for ordinal data, test statistics and model selection criteria have not yet been fully studied. This paper aims to derive likelihood ratio test statistics and model selection criteria under PML for SEM with ordinal variables. In particular, the mean-and-variance adjusted pairwise likelihood ratio test (PLRT) statistic for testing nested models and for testing overall goodness-of-fit together with their asymptotic distributions are derived. PLRT is the equivalent of the standard likelihood ratio test (LRT) under PML. Simulation examples study the performance of the proposed PLRT statistics for type I error and power and compare them to the mean-and-variance adjusted test statistics derived under the 3S-LS estimation methods. The performance of the pairwise likelihood model selection criteria, AIC_{PL} and BIC_{PL}, is also studied.

PML belongs to the family of composite likelihood (CL) estimation methods (Besag, 1974; Lindsay, 1988; Varin, 2008; Varin et al., 2011). The ML theory of inference has been extended to CL using the theory for misspecified likelihood functions. CL methods yield asymptotically consistent, and normally distributed estimators. Pace et al. (2011) present a Wald test, score test, and adjusted likelihood ratio test statistic for testing the hypothesis that a subset of parameters is equal to a specific value. Moreover, the model selection criteria AIC and the BIC are appropriately adjusted to hold under CL (Gao & Song, 2010; Varin et al., 2011; Varin & Vidoni, 2005). CL has gained attention because of its low computational complexity, which is not affected by model size. The advantage of CL is that it requires distributional assumptions about the lower-order margins and not for the complete variable vector as FIML does. Therefore, modelling assumptions are more straightforward, have less risk of misspecification, and are easier to test statistically. For example, Jöreskog (2002) discusses how the assumption of bivariate normality of two underlying continuous variables can be tested. The main argument against PML could be its loss of efficiency compared to FIML but simulation studies comparing the two methods, whenever FIML is practically feasible, indicate that this loss is minimal (Joe & Lee, 2009; Katsikatsou et al., 2012; Lele, 2006; Vasdekis et al., 2012; Zhao & Joe, 2005).

In SEM, De Leon (2005) proposes PML to estimate simultaneously the thresholds and polychoric correlations of ordinal variables. Liu (2007) extends the method to ordinal and continuous variables and proposes a two-stage estimation method in which thresholds and polychoric correlations are estimated using PML in the first stage, and the parameters of the factor model are estimated using generalised least squares in the second stage. The weight matrix is the PML estimate of the asymptotic covariance matrix of the estimated correlations. Furthermore, Liu (2007) derives a PML ratio test statistic for testing a hypothesis related to the parameters of the first stage (thresholds and polychoric correlations) and proposes a test statistic based on the generalised least squares fit function for testing the factor structure imposed on the polychoric correlations. Xi (2011), drawing on ideas from Jöreskog & Moustaki (2001), suggests a fit function composed of both the univariate and bivariate log-likelihood functions to fit a SEM. Xi (2011) notes that the test statistics developed under FIML cannot be directly applied under CL methods and proposes the implementation of a test statistic for overall fit based on bivariate residuals originally proposed by (Maydeu-Olivares & Joe, 2005, 2006). A pairwise likelihood estimation, where the likelihood function is defined as the product of the bivariate likelihoods, is proposed in Katsikatsou et al. (2012) for SEM for ordinal variables and in Katsikatsou (2013) for continuous and ranking data. PML estimation has been developed for panel models of ordered-responses (Bhat et al., 2010), latent variable models for ordinal longitudinal responses (Vasdekis et al., 2012), autoregressive ordered probit models (Varin & Vidoni, 2006), longitudinal mixed Rasch models (Feddag & Bacci, 2009), mixed models for joint modelling of multivariate longitudinal profiles (Fieuws & Verbeke, 2006), analysis of variance models (Lele & Taper, 2002), generalized linear models with crossed random effects (Bellio & Varin, 2005), spatial models with binary data (Heagerty & Lele, 1998), and spatial generalized linear mixed models (Varin et al., 2005) (see also the special issue of Statistica Sinica, Vol 21(1), 2011, for more areas of application).

The rest of the paper is organized as follows: Section 2 presents the SEM framework adopted here followed by a brief overview of the 3S-LS estimation and testing in Section 3. Section 4 describes the PML estimation for SEM and in Section 5, the formulae of PLRT statistics for overall goodness-of-fit and for testing nested models are derived. Section 6 provides the formulae of the model selection criteria AIC_{PL} and BIC_{PL}. Section 7 reports the results of the simulation study while Section 8 illustrates the proposed PLRT statistics using data from the European Social Survey. Conclusions and discussion are in Section 9. The proofs for the proposed test statistics are detailed in the Appendix and the R commands (R Development Core Team, 2008) used to obtain the presented results are given in the supplementary material. Our R code has been incorporated in the R package `lavaan` (Rosseel, 2012).

2 The Structural Equation Modelling framework

We follow the SEM framework discussed in Muthén (1984). Let $\mathbf{y}$ be an observed p-dimensional vector of ordinal variables. Let $\mathbf{y}^\star$ be the corresponding vector of underlying continuous variables. The connection between an ordinal variable y_i

and its underlying continuous variable $y_i^\star$ is: $y_i = a \Longleftrightarrow \tau_{i,a-1} < y_i^\star < \tau_{i,a}$, where a is the a-th response category of variable y_i, $a = 1, \ldots, c_i$, $i = 1, \ldots, p$, $\tau_{i,a}$ is the a-th threshold of variable y_i, and $-\infty = \tau_{i,0} < \tau_{i,1} < \ldots < \tau_{i,c_i-1} < \tau_{i,c_i} = +\infty$. Since only ordinal information is available, the distribution of $y_i^\star$ is determined only up to a monotonic transformation. It is typically assumed that $y_i^\star$ follows a standard normal distribution or a normal distribution with the mean and variance free to be estimated (e.g. Jöreskog, 2002). The measurement part of a SEM is:

$$\mathbf{y}^\star = \boldsymbol{\nu} + \Lambda\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{1}$$

and the structural part is:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathrm{B}\boldsymbol{\eta} + \boldsymbol{\zeta}\,, \tag{2}$$

where $\boldsymbol{\eta}$ is a q-dimensional vector of continuous latent variables, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}$ are the vectors of error terms, and $\boldsymbol{\nu}$ and $\boldsymbol{\alpha}$ are the vectors of intercepts.

The standard basic assumptions of the model are that: $\mathbf{y}^\star \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\eta}$ follows a multivariate normal distribution, $\boldsymbol{\varepsilon} \sim \mathcal{N}_p(\mathbf{0}, \Theta)$, $\boldsymbol{\zeta} \sim \mathcal{N}_q(\mathbf{0}, \Psi)$, $Cov(\boldsymbol{\eta}, \boldsymbol{\varepsilon}) = Cov(\boldsymbol{\eta}, \boldsymbol{\zeta}) = Cov(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}) = \mathbf{0}$, and $I - \mathrm{B}$ is non-singular with I being the identity matrix. From (2), it follows that $E(\boldsymbol{\eta}) = (I - \mathrm{B})^{-1}\boldsymbol{\alpha}$ and $Cov(\boldsymbol{\eta}) = (I - \mathrm{B})^{-1}\Psi\left[(I - \mathrm{B})^{-1}\right]'$. Thus, the model-implied mean vector $\boldsymbol{\mu}$ and covariance matrix Σ of $\mathbf{y}^\star$ are:

$$\boldsymbol{\mu} = E(\mathbf{y}^\star) = \boldsymbol{\nu} + \Lambda(I - \mathrm{B})^{-1}\boldsymbol{\alpha}\,,$$

$$\Sigma = Cov(\mathbf{y}^\star) = \Lambda(I - \mathrm{B})^{-1}\Psi\left[(I - \mathrm{B})^{-1}\right]'\Lambda' + \Theta\,.$$

Depending on the specific model, further constraints including those for identification may be required. The scale of all underlying variables $\mathbf{y}^\star$ and the latent variables need to be defined. In the case of multi-group analysis, a minimum set of restrictions is needed so that the model is identified and a common scale for each latent variable is defined across groups (Millsap & Yun-Tein, 2004; Muthén & Asparouhov, 2002).

3 Three-stage least squares approach

Under a 3S-LS estimation, in the first stage, first order statistics such as thresholds, means and variances are estimated by maximum likelihood. In the second stage, second order statistics such as polychoric correlations are estimated by conditional maximum likelihood for given first stage estimates. In the third stage, the parameters of the structural part of the model are estimated using a generalized or weighted least squares method. The fit function to be minimized is of the form:

$$F(\boldsymbol{\theta}) = (\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta}))'\,W^{-1}(\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta}))\,, \tag{3}$$

where $\mathbf{r}$ is the vector of sample statistics (e.g. thresholds, polychoric correlations), $\boldsymbol{\rho}$ is the vector of their model-implied counterparts, and $\boldsymbol{\theta}$ is the model parameter vector. The weight matrix W is either the estimated asymptotic covariance matrix of the sample statistics (weighted least squares (WLS)), or a diagonal matrix (diagonally weighted least squares (DWLS)), or the identity matrix (unweighted least squares (ULS)). Under all three estimation methods (WLS, DWLS, ULS), the full estimated asymptotic covariance matrix is used to compute the standard errors and goodness-of-fit test statistics.

Under both DWLS and ULS, the test statistic for overall fit is written as $T = (N-1)\,F\left(\hat{\boldsymbol{\theta}}\right)$, where F is the fit function in Equation (3) evaluated at $\hat{\boldsymbol{\theta}}$ and N is the sample size. Various adjusted versions of T have been proposed in the literature (Asparouhov & Muthén, 2010; Muthén, 1993; Muthén et al., 1997; Satorra & Bentler, 1994). Savalei & Rhemtulla (2013) compare the different versions of the test statistics through an extensive simulation study. They found that the mean-and-variance adjusted T following the Satterthwaite approximation has the best performance in terms of type I error and power. The exact formulae of mean-and-variance adjusted T derived under DWLS, $T_{DWLS-MV}$, and under ULS, T_{ULS-MV}, are provided in Equations (2) and (3) of their paper, respectively.

For the testing of nested models under the 3S-LS methods, Satorra (2000) proposes a test statistic given by the difference of the estimated fit functions adjusted in mean and variance using the Satterthwaite approximation. The obtained test statistic is asymptotically chi-squared distributed. Asparouhov & Muthén (2006) show that this statistic works well for categorical data too. The same statistic, but only adjusted in mean, has also been discussed by Satorra & Bentler (2001); Asparouhov & Muthén (2006); Satorra & Bentler (2010). However, it is well known that mean-and-variance adjusted chi-squared statistics perform better in smaller sample sizes and converge faster to their asymptotic properties than the corresponding mean-adjusted ones.

4 Pairwise likelihood estimation

The PML function to be maximized for estimating a factor analysis model with ordinal variables is given in Katsikatsou et al. (2012). Let $\boldsymbol{\theta}$ be the parameter vector that includes the free thresholds and parameters: $\boldsymbol{\nu}$, $\boldsymbol{\alpha}$, Λ, B, Γ, Ψ, and Θ defined in Section 2. For a random sample of N observations the pairwise log-likelihood (pl) is defined as follows:

$$pl\left(\boldsymbol{\theta};\mathbf{y}\right) = pl\left(\boldsymbol{\theta};\left(\mathbf{y}_1,\ldots,\mathbf{y}_N\right)\right) = \sum_{n=1}^{N}\sum_{i<i'}\ln L\left(\boldsymbol{\theta};\left(y_{in},y_{i'n}\right)\right) . \tag{4}$$

The specific form of the bivariate log-likelihood $\ln L\left(\boldsymbol{\theta};(y_{in}, y_{i'n})\right)$ for a single observation is:

$$\ln L\left(\boldsymbol{\theta};(y_i, y_{i'})\right) = \sum_{a=1}^{c_i}\sum_{a'=1}^{c_{i'}} I\left(y_i = a, y_{i'} = a'\right) \ln \pi\left(y_i = a, y_{i'} = a'; \boldsymbol{\theta}\right) ,$$

where $I\left(y_i = a, y_{i'} = a'\right)$ is an indicator variable taking the value 1 if y_i and $y_{i'}$ fall into categories a and a', respectively, and 0 otherwise,

$$\pi\left(y_i = a, y_{i'} = a'; \boldsymbol{\theta}\right) = \int_{\tau_{i,a-1}}^{\tau_{i,a}} \int_{\tau_{i',a'-1}}^{\tau_{i',a'}} f\left(y_i^\star, y_{i'}^\star\right) dy_i^\star dy_{i'}^\star , \tag{5}$$

and $f\left(y_i^\star, y_{i'}^\star\right)$ is the density of the corresponding underlying variables $y_i^\star$ and $y_{i'}^\star$ taken to be a bivariate normal distribution with mean vector $(\mu_i, \mu_{i'})'$ and covariance matrix with elements: $\sigma_{ii}, \sigma_{ii'}, \sigma_{i'i'}$. The means, the variances, and the covariances of the underlying variables are functions of the parameter vector $\boldsymbol{\theta}$. The value of $\boldsymbol{\theta}$ that maximizes the *pl* function given the data at hand (Equation (4)) is defined to be the PML estimator, $\hat{\boldsymbol{\theta}}_{PL}$. Since PML estimation assumes bivariate normality for all pairs of variables in $\mathbf{y}^\star$ it requires the evaluation of two-dimensional normal probabilities (Equation (5)) regardless of the number of observed variables. In practice, the maximization is carried out numerically and for this the analytical form of the gradient of the *pl* function is required (given in Sections A.2. and A.3. in Katsikatsou, 2013).

From the theory of CL estimators, it holds that $\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{PL} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}\left(0, G^{-1}(\boldsymbol{\theta})\right),$ where $G(\boldsymbol{\theta})$ is the Godambe information matrix (also known as the sandwich information matrix), $G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J^{-1}(\boldsymbol{\theta})H(\boldsymbol{\theta})$, $H(\boldsymbol{\theta}) = E\left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}} pl(\boldsymbol{\theta};\mathbf{y})\right\}$, and $J(\boldsymbol{\theta}) = Var\left\{\frac{\partial}{\partial\boldsymbol{\theta}'} pl(\boldsymbol{\theta};\mathbf{y})\right\}$ (Lindsay, 1988; Varin et al., 2011). In general, the identity $H(\boldsymbol{\theta}) = -J(\boldsymbol{\theta})$ does not hold under CL because the assumed independence among the likelihood components forming the CL is not valid when the full likelihood is considered. $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ can be estimated by:

$$\hat{H}(\hat{\boldsymbol{\theta}}_{PL}) = -\frac{1}{N}\frac{\partial^2}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}} pl\left(\boldsymbol{\theta};(\mathbf{y}_1,\ldots,\mathbf{y}_{N_g})\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{PL}} , \tag{6}$$

$$\hat{J}(\hat{\boldsymbol{\theta}}_{PL}) = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\partial}{\partial\boldsymbol{\theta}'} pl\left(\boldsymbol{\theta};\mathbf{y}_n\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{PL}}\right)\left(\frac{\partial}{\partial\boldsymbol{\theta}'} pl\left(\boldsymbol{\theta};\mathbf{y}_n\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{PL}}\right)' . \tag{7}$$

5 Pairwise likelihood ratio test statistic

The pairwise likelihood ratio test is derived under PML estimation for testing the overall fit of a model and for comparing nested models. We show that asymptotically the PLRT statistic, both for the overall fit and for testing nested models, is a weighted sum of independent chi-squared variables. To determine the asymptotic distribution of PLRT the Satterthwaite approximation is used which leads to the mean-and-variance adjusted PLRT. This requires computing the asymptotic mean and variance of the test statistic under H_0. The proofs, given in Appendices A.1 and A.2, use Taylor series expansions, the asymptotic normality of the pairwise likelihood estimator, and the standard assumption that the null hypothesis is true.

5.1 Pairwise likelihood ratio test statistic for nested models

Let $\boldsymbol{\theta}$ be the parameter vector of dimension d under H_1 and $g(\boldsymbol{\theta})$ be a function of $\boldsymbol{\theta}$, where $g : \mathbb{R}^d \to \mathbb{R}^r$, and r is the number of constraints. Let the hypothesis of interest be $H_0 : g(\boldsymbol{\theta}) = \mathbf{0}$ versus $H_1 : g(\boldsymbol{\theta}) \neq \mathbf{0}$. The PLRT statistic is

$$PLRT\left(g\left(\boldsymbol{\theta}\right)\right) = 2\left(pl\left(\hat{\boldsymbol{\theta}}\right) - pl\left(\tilde{\boldsymbol{\theta}}\right)\right) \,, \tag{8}$$

where $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ are the PML estimates under H_1 and H_0, respectively. Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$. It can be shown (the proof is given in Appendix A.1) that:

$$PLRT\left(g\left(\boldsymbol{\theta}\right)\right) \;\; \to \;\; \tilde{\boldsymbol{z}}'\tilde{\boldsymbol{z}} \,,$$

where $\tilde{\boldsymbol{z}} = \sqrt{N}[A(\boldsymbol{\theta}_0)]^{-1/2} g\left(\hat{\boldsymbol{\theta}}\right)$, $\sqrt{N} g\left(\hat{\boldsymbol{\theta}}\right) \to \mathcal{N}\left(\mathbf{0}, B(\boldsymbol{\theta}_0)\right)$,
$A(\boldsymbol{\theta}_0) = M(\boldsymbol{\theta}_0) H^{-1}(\boldsymbol{\theta}_0) [M(\boldsymbol{\theta}_0)]'$, $B(\boldsymbol{\theta}_0) = M(\boldsymbol{\theta}_0) G^{-1}(\boldsymbol{\theta}_0) [M(\boldsymbol{\theta}_0)]'$, and
$M(\boldsymbol{\theta}_0) = \frac{\partial}{\partial \boldsymbol{\theta}'} g(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ is an $r \times d$ matrix of the gradient of function g with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_0$. Hence, $\tilde{\boldsymbol{z}} \to \mathcal{N}\left(\mathbf{0}, [A(\boldsymbol{\theta}_0)]^{-1/2} B(\boldsymbol{\theta}_0)\{[A(\boldsymbol{\theta}_0)]^{-1/2}\}'\right)$ and $PLRT\left(g\left(\boldsymbol{\theta}\right)\right) \to \sum_{i=1}^{r} \kappa_i u_i$, where κ_i is the ith eigenvalue of $[A(\boldsymbol{\theta}_0)]^{-1/2} B(\boldsymbol{\theta}_0)\{[A(\boldsymbol{\theta}_0)]^{-1/2}\}'$ and u_i's are independent χ^2_1-distributed variables. To determine the asymptotic distribution of $PLRT\left(g\left(\boldsymbol{\theta}\right)\right)$ we apply the Satterthwaite approximation. Under H_0, the asymptotic mean and variance of $PLRT\left(g\left(\boldsymbol{\theta}\right)\right)$ are:

$$E\left[PLRT\left(g\left(\boldsymbol{\theta}\right)\right)\right] \to \operatorname{tr}\left(B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\right) \,, \text{ and} \tag{9}$$

$$Var\left[PLRT\left(g\left(\boldsymbol{\theta}\right)\right)\right] \to 2\operatorname{tr}\left(B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1} B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\right) \,. \tag{10}$$

Let $PLRT_{MV}\left(g\left(\boldsymbol{\theta}\right)\right)$ denote the mean-and-variance adjusted $PLRT\left(g\left(\boldsymbol{\theta}\right)\right)$. Under H_0, it holds that:

$$PLRT_{MV}\left(g\left(\boldsymbol{\theta}\right)\right) = \alpha\left(\boldsymbol{\theta}_0\right) PLRT\left(g\left(\boldsymbol{\theta}\right)\right) \overset{\text{app}}{\to} \chi^2_{df(\boldsymbol{\theta}_0)} \,,$$

where $\alpha(\boldsymbol{\theta}_0) = \frac{\operatorname{tr}\left(B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\right)}{\operatorname{tr}(B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1} B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1})}$ and $df(\boldsymbol{\theta}_0) = \frac{\left[\operatorname{tr}\left(B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1}\right)\right]^2}{\operatorname{tr}(B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1} B(\boldsymbol{\theta}_0)[A(\boldsymbol{\theta}_0)]^{-1})}$.
In practice, since $\boldsymbol{\theta}_0$ is unknown, $\alpha\left(\tilde{\boldsymbol{\theta}}\right)$ and $df\left(\tilde{\boldsymbol{\theta}}\right)$ are used instead. This is why the degrees of freedom in the application will be subject to sample variability.

A special case is the hypothesis $H_0 : \boldsymbol{\psi} = \boldsymbol{\psi}_0$ versus $H_1 : \boldsymbol{\psi} \neq \boldsymbol{\psi}_0$, where $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\omega}')'$, $\boldsymbol{\psi}$ is the vector of parameters of interest, $\boldsymbol{\omega}$ is the vector of nuisance parameters, and $\boldsymbol{\psi}_0$ is a vector of real values. Then, the results for the asymptotic mean and variance of PLRT given in expressions (9) and (10) simplify to:

$$E\left[PLRT(\boldsymbol{\psi})\right] \to \mathrm{tr}\left(G^{\psi\psi}\left(\boldsymbol{\theta}_0\right)\left[H^{\psi\psi}\left(\boldsymbol{\theta}_0\right)\right]^{-1}\right), \text{ and}$$

$$Var\left[PLRT(\boldsymbol{\psi})\right] \to 2\mathrm{tr}\left(G^{\psi\psi}\left(\boldsymbol{\theta}_0\right)\left[H^{\psi\psi}\left(\boldsymbol{\theta}_0\right)\right]^{-1} G^{\psi\psi}\left(\boldsymbol{\theta}_0\right)\left[H^{\psi\psi}\left(\boldsymbol{\theta}_0\right)\right]^{-1}\right),$$

where $G^{\psi\psi}\left(\boldsymbol{\theta}_0\right)$ and $H^{\psi\psi}\left(\boldsymbol{\theta}_0\right)$ are, respectively, the parts of the inverse of $G\left(\boldsymbol{\theta}_0\right)$ and $H\left(\boldsymbol{\theta}_0\right)$ matrices that refer to the parameter vector $\boldsymbol{\psi}$. The simplification occurs because the matrix $M\left(\boldsymbol{\theta}_0\right)$ becomes an indicator matrix that consists of 0's and only one 1 in each row where the 1's are in the columns that correspond to the parameters constrained under H_0. The role of matrix $M\left(\boldsymbol{\theta}_0\right)$ in the calculation of matrices $B(\boldsymbol{\theta}_0)$ and $A\left(\boldsymbol{\theta}_0\right)$ is to pick the right parts of $G^{-1}\left(\boldsymbol{\theta}_0\right)$ and $H^{-1}\left(\boldsymbol{\theta}_0\right)$, respectively.

The proposed $PLRT_{MV}\left(g\left(\boldsymbol{\theta}\right)\right)$ statistic for the hypothesis $H_0 : g\left(\boldsymbol{\theta}\right) = \mathbf{0}$ versus $H_1 : g\left(\boldsymbol{\theta}\right) \neq \mathbf{0}$ holds when $g\left(\boldsymbol{\theta}\right)$ includes both equality constraints among parameters and constraints where some parameters are set equal to specific values.

5.2 Pairwise likelihood ratio test statistic for overall fit

We first consider the case where a model imposes a parametric structure on the covariance matrix Σ and not on thresholds. Let $\boldsymbol{\varphi}$ be a d-dimensional vector of all model parameters but the thresholds. Let $\boldsymbol{\tau}$ be the vector of thresholds. Let $\boldsymbol{\theta}$ be the complete parameter vector, thus, $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\tau}')'$. Let $\boldsymbol{\sigma} = \mathrm{vech}\left(\Sigma\right)$, where vech is the vectorization function of the elements of Σ being on and below the main diagonal, and $\boldsymbol{\sigma}$ is of dimension $\tilde{p}$ which is the number of free non-redundant elements of Σ. The null hypothesis for overall model fit is written as $H_0 : \boldsymbol{\sigma} = g(\boldsymbol{\varphi})$ versus $H_1 : \boldsymbol{\sigma}$ unconstrained, where g is a model-dependent function, $g : \mathbb{R}^d \to \mathbb{R}^{\tilde{p}}$. Note that H_0 does not include the threshold vector $\boldsymbol{\tau}$, hence, it is a nuisance parameter. Under H_0, it holds that $pl\left(\boldsymbol{\theta}\right) = pl\left(\boldsymbol{\vartheta}\right)$, where $\boldsymbol{\vartheta}$ is the complete parameter vector under H_1, and $\boldsymbol{\vartheta} = (\boldsymbol{\sigma}', \boldsymbol{\tau}')'$. If $\boldsymbol{\theta}_0 = (\boldsymbol{\varphi}_0', \boldsymbol{\tau}_0')'$ is the true value of the parameter, then $\boldsymbol{\vartheta}_0 = \left(g\left(\boldsymbol{\varphi}_0\right)', \boldsymbol{\tau}_0'\right)' = (\boldsymbol{\sigma}_0', \boldsymbol{\tau}_0')'$. The PLRT statistic is defined as before:

$$PLRT_{SEM} = 2\left(pl\left(\hat{\boldsymbol{\vartheta}}\right) - pl\left(\hat{\boldsymbol{\theta}}\right)\right), \tag{11}$$

where $\hat{\boldsymbol{\vartheta}} = (\hat{\boldsymbol{\sigma}}', \hat{\boldsymbol{\tau}}')'$ and $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\tau}}')'$ are the PML estimates under H_1 and H_0, respectively. It can be shown that under H_0 (the proof is given in Appendix A.2.):

$$PLRT_{SEM} \to \boldsymbol{z}'\boldsymbol{z} - \boldsymbol{v}'\boldsymbol{v}\,, \tag{12}$$

where $\boldsymbol{z} = \sqrt{N}\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1/2}\left(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0\right)$, $\boldsymbol{v} = \sqrt{N}\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1/2}\left(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0\right)$,

$$\boldsymbol{z} \rightarrow \mathcal{N}_{\tilde{p}}\left(\mathbf{0}, \left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1/2} G^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1/2}\right), \text{ and} \tag{13}$$

$$\boldsymbol{v} \rightarrow \mathcal{N}_{d}\left(\mathbf{0}, \left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1/2} G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1/2}\right). \tag{14}$$

The matrices $H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)$, $G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)$, $H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)$, and $G^{\sigma\sigma}(\boldsymbol{\vartheta}_0)$ are defined similarly to $H^{\psi\psi}(\boldsymbol{\theta}_0)$ and $G^{\psi\psi}\left(\boldsymbol{\theta}_0\right)$ above. From (12), (13), and (14) it follows that $PLRT_{SEM}$ is asymptotically the difference of two weighted sums of independent chi-squared variables. To apply the Satterthwaite approximation we compute the asymptotic mean and variance of $PLRT_{SEM}$ given by:

$$E\left(PLRT_{SEM}\right) \rightarrow \operatorname{tr}\left(G^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1}\right) - \operatorname{tr}\left(G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1}\right), \tag{15}$$

$$\begin{aligned} Var\left(PLRT_{SEM}\right) \rightarrow\ & 2\operatorname{tr}\left(G^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1} G^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1}\right) \\ & + 2\operatorname{tr}\left(G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1} G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1}\right) \\ & - 4\operatorname{tr}\left(M'\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1} M G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1} G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right), \end{aligned} \tag{16}$$

where $M = \frac{\partial}{\partial \boldsymbol{\varphi}} g\left(\boldsymbol{\varphi}\right)\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_0}$. The computation of the asymptotic $Var\left(PLRT_{SEM}\right)$ is given in Appendix A.3. Let $\alpha_1\left(\boldsymbol{\theta}_0\right)$ and $\alpha_2\left(\boldsymbol{\theta}_0\right)$ denote the right hand side of expressions (15) and (16), respectively. Let $PLRT_{SEM-MV}$ denote the mean-and-variance adjusted $PLRT_{SEM}$. Under H_0, it holds that:

$$PLRT_{SEM-MV} = \alpha\left(\boldsymbol{\theta}_0\right) PLRT_{SEM} \overset{\text{app}}{\rightarrow} \chi^2_{df(\boldsymbol{\theta}_0)},$$

where $\alpha\left(\boldsymbol{\theta}_0\right) = \frac{\alpha_1(\boldsymbol{\theta}_0)}{0.5 * \alpha_2(\boldsymbol{\theta}_0)}$, $df\left(\boldsymbol{\theta}_0\right) = \frac{[\alpha_1(\boldsymbol{\theta}_0)]^2}{0.5 * \alpha_2(\boldsymbol{\theta}_0)}$. Observe that, as before, both the adjustment coefficient $\alpha\left(\boldsymbol{\theta}_0\right)$ and the adjusted degrees of freedom $df\left(\boldsymbol{\theta}_0\right)$ are functions of the true value $\boldsymbol{\theta}_0$ which, in practice, is substituted by its PML estimate under H_0, $\hat{\boldsymbol{\theta}}$. Hence, both quantities are subject to sample variability.

In the case of a model which imposes a parametric structure both on the covariance matrix Σ and on thresholds, the hypothesis is modified to $H_0 : \boldsymbol{\vartheta} = g(\boldsymbol{\theta})$ versus $H_1 : \boldsymbol{\vartheta}$ unconstrained. All the above results remain the same with the only difference being that in expressions (15) and (16), $G^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)$, $\left[H^{\sigma\sigma}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1}$, $G^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)$, and $\left[H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)\right]^{-1}$ are substituted with $G^{-1}\left(\boldsymbol{\vartheta}_0\right)$, $H\left(\boldsymbol{\vartheta}_0\right)$, $G^{-1}\left(\boldsymbol{\theta}_0\right)$, and $H\left(\boldsymbol{\theta}_0\right)$, respectively, and $M = \frac{\partial}{\partial \boldsymbol{\theta}} g\left(\boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

The PLRT for overall fit is of the same nature as the test statistics derived under 3S-LS in the sense that the parametric structure imposed by the model on the thresholds and the covariance matrix is being tested.

6 Pairwise likelihood model selection criteria

This section discusses the AIC and BIC model selection criteria for SEM under PML estimation. Based on the results of Varin & Vidoni (2005), the Akaike PML information criterion, AIC_{PL}, is defined as:

$$AIC_{PL} = -pl\left(\hat{\boldsymbol{\theta}}_{PL};\mathbf{y}\right) + tr(\hat{J}(\hat{\boldsymbol{\theta}}_{PL})\hat{H}^{-1}(\hat{\boldsymbol{\theta}}_{PL})), \tag{17}$$

and, based on the results of Gao & Song (2010), the PML Bayesian information criterion, BIC_{PL}, is defined as:

$$BIC_{PL} = -2pl\left(\hat{\boldsymbol{\theta}}_{PL};\mathbf{y}\right) + tr(\hat{J}(\hat{\boldsymbol{\theta}}_{PL})\hat{H}^{-1}(\hat{\boldsymbol{\theta}}_{PL})) \times \log N\,, \tag{18}$$

where $\hat{\boldsymbol{\theta}}_{PL}$ is the PML estimate under the hypothesized model, and $tr(\hat{J}(\hat{\boldsymbol{\theta}}_{PL})\hat{H}^{-1}(\hat{\boldsymbol{\theta}}_{PL}))$ defines the number of effective parameters. The model with the smallest AIC_{PL} or BIC_{PL} is selected.

7 Simulation study

The type I error and power of the proposed mean-and-variance PLRT statistics for overall fit and for testing nested models are assessed using simulations studies. The data were simulated on the basis of combinations of sample size, number of response categories, and model complexity.

The empirical rejection rates of the null hypothesis are computed as follows: let $t^{(r)}$ and $df^{(r)}$ be the rth replicated values of a test statistic and its associated estimated degrees of freedom. Then, the p-value from the rth replication is $p\text{-value}^{(r)} = \Pr\left(w > t^{(r)}\right)$ where $w \sim \chi^2_{df^{(r)}}$ and the rejection rate is the percentage of $p\text{-value}^{(r)}$'s out of the total replications that are smaller than or equal to the nominal significance level 5% and 1%. Note that in each replication, the adjustment coefficient $\alpha\left(\boldsymbol{\theta}_0\right)$ and the adjusted degrees of freedom $df\left(\boldsymbol{\theta}_0\right)$ are computed by substituting $\boldsymbol{\theta}_0$ with the rth replicated PML estimate under H_0, $\hat{\boldsymbol{\theta}}^{(r)}_{PL}$, and by using the sample estimates of $H(\boldsymbol{\theta})$ and $J(\boldsymbol{\theta})$ matrices given in expressions (6) and (7), respectively. The computation of these sample estimates involves the complete rth replicated sample. The sample estimate of $J(\boldsymbol{\theta})$ is preferred here to the theoretical one as the latter is complicated to compute. Also, the use of the observed information matrix has been often proposed against the expected information matrix (e.g. Efron & Hinkley, 1978; Kenward & Molenberghs, 1998).

The performance of PLRT is also compared with that of the corresponding test statistics derived under DWLS and ULS, $T_{DWLS-MV}$ and T_{ULS-MV}. For overall fit, we compute the formulae of $T_{DWLS-MV}$ and T_{ULS-MV} given in expressions

(2) and (3) in Savalei & Rhemtulla (2013), respectively, and for comparing nested models, the formulae given in Satorra (2000) (page 243, end of Section 3). The performance of AIC_{PL} and BIC_{PL} is also studied. For all computations including those under the PML method, we use the R package `lavaan`.

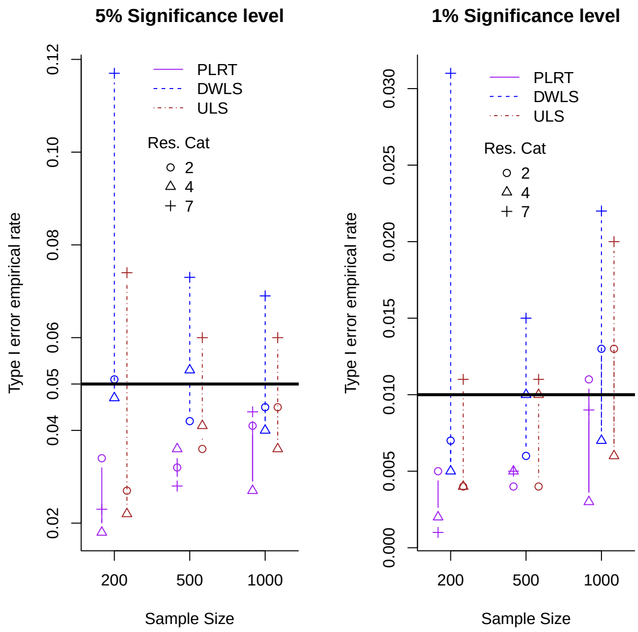
7.1 On the performance of PLRT for overall fit

The performance of $PLRT_{SEM-MV}$ for overall fit is studied for type I error and power. For type I error, nine experimental conditions are considered. We study three sample sizes, 200, 500, and 1000, and three different numbers of response categories namely two, four, and seven. Within each experimental condition, 1000 replications are carried out. The data are generated by a confirmatory two-factor model with 20 ordinal variables where each factor is measured by 10 indicators (Model 0). The loadings of each set of variables are 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.8, and 0.9. The correlation between the two factors is 0.4. The values of the thresholds are: 0 when the indicators are binary; -1.25, 0, and 1.25 when they have four response categories; and -1.79, -1.07, -0.36, 0.36, 1.07, 1.79 when they have seven response categories. This way, the theoretical distribution of each ordinal variable is assumed to be symmetric.

For all conditions, except for sample size 200 and 2 response categories, all three methods (PL, DWLS, ULS) show 100% convergence rate and 100% rate of proper solutions (i.e. all estimated variances are positive and all correlations are between -1 and 1). For sample size 200 and 2 response categories, despite the convergence rate being 100% for all three methods, the rate of proper solutions is 97.8% for PML and DWLS and 94.9% for ULS. The results regarding the test statistics reported below are based on the total number of replications because the full output is produced for all of them and improper solutions are expected to happen in small sample sizes and do not necessarily represent a statistical anomaly (Savalei & Kolenikov, 2008; Savalei & Rhemtulla, 2013).

Figure 1 gives the empirical type I error rates for each method and experimental condition. In each subfigure, the bold horizontal line represents the nominal significance level set at 5% and 1%. The empirical Type I rates for the $PLRT_{SEM-MV}$ are satisfactory for half of the experimental conditions studied, mainly when the sample size is larger and the nominal significance level is 1%. The number of response categories do not seem to have a clear effect on the empirical rates. It is noted that whenever $PLRT_{SEM-MV}$ fails to reach the nominal significance level, it under-rejects the null hypothesis. The performance of $T_{DWLS-MV}$ and T_{ULS-MV} is slightly better than the $PLRT_{SEM-MV}$, except for the case of 7 response categories where both statistics over-reject the model. The performance of $T_{DWLS-MV}$ does not seem to improve with the increase in sample size and is particularly unsatisfactory for sample size 200. Similar results about $T_{DWLS-MV}$ and T_{ULS-MV}

Figure 1: Empirical type I error rates for the three overall-fit test statistics, $\mathrm{PLRT}_{\mathrm{SEM-MV}}$, $\mathrm{T}_{\mathrm{DWLS-MV}}$, $\mathrm{T}_{\mathrm{ULS-MV}}$, for data with 2, 4 and 7 response categories and samples sizes 200, 500 and 1000; the bold horizontal lines represent the nominal significance level; the vertical lines joining the symbols (circle, triangle, cross) are used to distinguish among the three test statistics and do not represent a range of values



are reported in Savalei & Rhemtulla (2013).

The empirical type I error rates along with their 95% confidence interval for the three test statistics, and the average of the replicated degrees of freedom for each method and experimental condition are reported in Table 1 that can be found in the supplementary material. The medians of the replicated degrees of freedom are not reported because they are found to be very close to the corresponding means in all experimental conditions (absolute differences less than 0.6). The Q-Q plots for $PLRT_{SEM-MV}$ for all nine experimental conditions are also provided in the supplementary material. In these plots the interest lies on the higher quantiles (for example, 90% or higher) as $PLRT_{SEM-MV}$ is a test statistic for overall fit.

The power for the three test statistics for overall fit is investigated under three model misspecifications. Under misspecifications 1 and 2, the fitted model is similar to the data-generating model (Model 0) with the only difference that the factor correlation is fixed to 0.3 (Model 1a) and 0 (Model 1b), respectively. The experimental conditions remain the same as above. Under misspecification 3, the data generating model is a confirmatory two-factor model in which variables 1 to 10 load on the first factor with corresponding loadings 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.8, 0.8, while variables 8 to 20 load on the second factor with corresponding loadings 0.2, 0.2, 0.2, 0.3, 0.4, 0.4, 0.5, 0.5, 0.6, 0.6, 0.7, 0.8, 0.9. The factor correlation is set to 0.4, and all variables have four response categories with the thresholds being equal to -1.25, 0, 1.25. The fitted model misspecifies the loadings on the second factor for variables 8-10 by fixing them to zero. Three sample sizes, 200, 500, 1000, are considered.

The convergence rate is 100% for all three methods and all simulation conditions. The rate of proper solutions is 100% except for the case of 2 response categories and 200 sample size, where the rates for PML, DWLS, and ULS, respectively, are 96.7%, 96%, and 89% when Model 1a is fitted; and 98.5%, 98.5%, and 97.5% when Model 1b is fitted. In addition to this, the ULS rate of proper solution, when Model 1a is fitted, is: 98.8% for 2 response categories and 500 sample size, 98.9% for 4 response categories and 200 sample size, and 99.6% for 7 response categories and 200 sample size. Moreover, under Misspecification 3, the rate for ULS is 99% and 99.9% for sample sizes 200 and 500, respectively.

Figure 2 and Table 2 (in the supplementary material) show the results for Misspecification 1. For all three test statistics, the power increases with the sample size and with the number of response categories at both nominal significance levels. In all experimental conditions, $T_{DWLS-MV}$ and T_{ULS-MV} perform slightly better than $PLRT_{SEM-MV}$ but the differences decrease as the sample size increases. The slightly lower power of $PLRT_{SEM-MV}$ is expected as it tends to under-reject a true null hypothesis. Figure 3 and Table 3 (in the supplementary material) show the results for Misspecification 2. For this larger misspecification, the power of all three statistics is close to 1 for sample size 200 and is exactly 1 for sample size 500 for all three different numbers of response categories. For sample size 200, the differences among the three test statistics are negligible.

Figure 4 and Table 4 (in the supplementary material) and show the results under Misspecification 3. The power for all three test statistics is rather low for sample size 200 but improves substantially with the increase of sample size. It gets close to 1 for sample size 1000. Among the three test statistics, $T_{DWLS-MV}$ performs slightly better, while T_{ULS-MV} and $PLRT_{SEM-MV}$ perform similarly. The differences become negligible as the sample size increases.

7.2 On the performance of PLRT, AIC_{PL}, and BIC_{PL} for nested models

The performance of $PLRT_{MV}$, AIC_{PL}, and BIC_{PL} for nested models with respect to type I error is studied under two different settings: a) in a single-group

Figure 2: Empirical power rates for the overall-fit test statistics, $\mathrm{PLRT}_{\mathrm{SEM-MV}}$, $\mathrm{T}_{\mathrm{DWLS-MV}}$, $\mathrm{T}_{\mathrm{ULS-MV}}$, for data with 2, 4, and 7 response categories, sample sizes 200, 500, 1000, and nominal significance levels 5% and 1%; the fitted model (Model 1a) misspecifies the factor correlation by fixing it equal to 0.3 while the true value is 0.4; the vertical lines joining the symbols (circle, triangle, cross) are used to distinguish among the three test statistics and do not represent a range of values

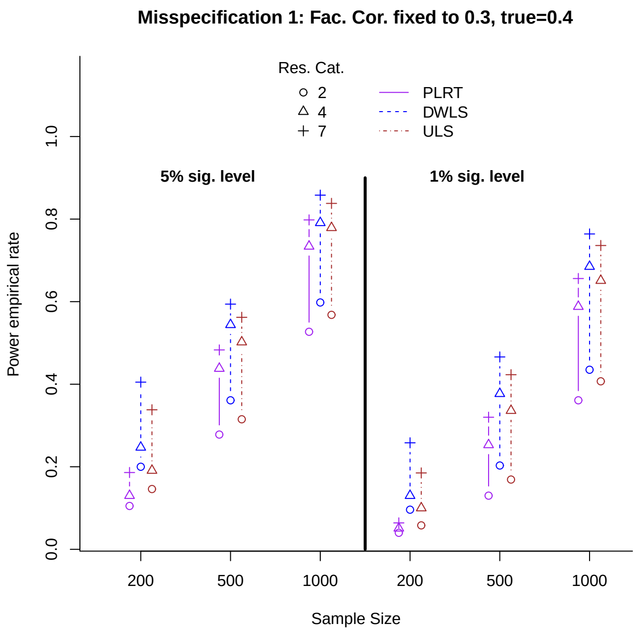


Figure 3: Empirical power rates for the overall-fit test statistics, $\text{PLRT}_{\text{SEM-MV}}$, $\text{T}_{\text{DWLS-MV}}$, $\text{T}_{\text{ULS-MV}}$, for data with 2, 4, and 7 response categories, sample sizes 200, 500, and nominal significance levels 5% and 1%; the fitted model (Model 1b) misspecifies the factor correlation by fixing it equal to 0 while the true value is 0.4; the vertical lines joining the symbols (circle, triangle, cross) are used to distinguish among the three test statistics and do not represent a range of values

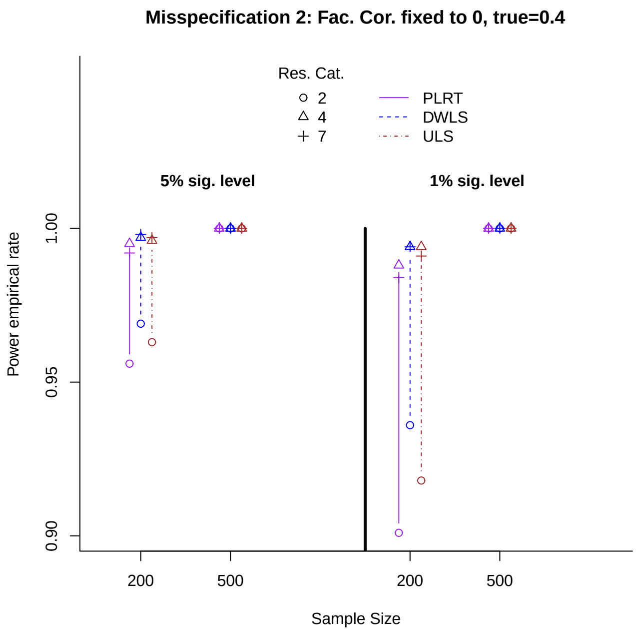
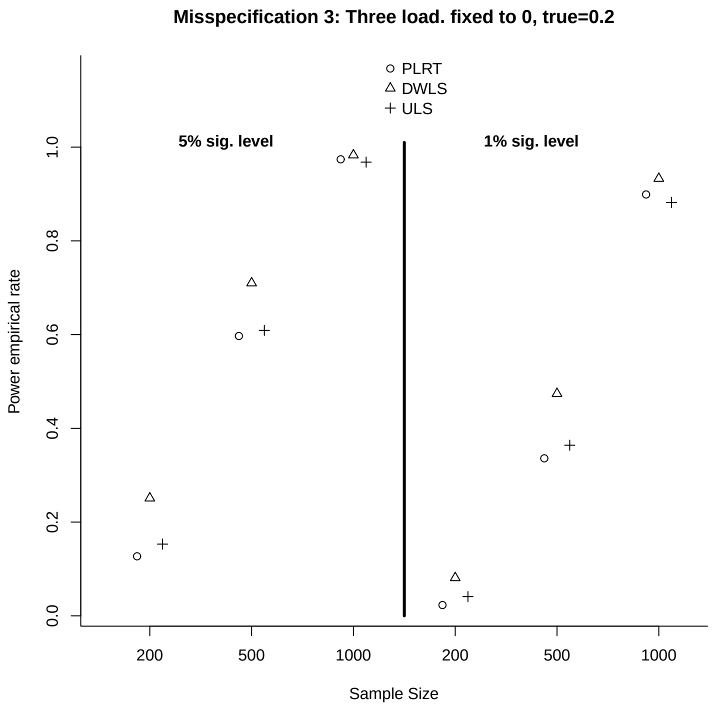


Figure 4: Empirical power rates for the overall-fit test statistics, $\mathrm{PLRT}_{\mathrm{SEM-MV}}$, $\mathrm{T}_{\mathrm{DWLS-MV}}$, $\mathrm{T}_{\mathrm{ULS-MV}}$, for data with 4 response categories, sample sizes 200, 500, 1000, and significance levels 5% and 1%; the fitted model (Model 0) misspecifies three loadings by fixing them equal to 0 while their true value is 0.2



analysis where models are nested due to parameter constraints (some parameters are set equal to zero), and b) in a multi-group analysis where measurement equivalence across groups translates statistically into a series of comparisons of nested models due to cross-group equality constraints on the measurement model parameters. In particular, the model in Equations (1) and (2) can be extended to multi-group analysis by adding a superscript g to all variables and parameters with g denoting the group membership, $g = 1, \ldots, G$, and G is the number of independent groups. This way, the pl log-likelihood in Equation (4) is modified to $pl\left(\boldsymbol{\theta}; \mathbf{y}\right) = \sum_{g=1}^{G} pl\left(\boldsymbol{\theta}; \left(\mathbf{y}_1^{(g)}, \ldots, \mathbf{y}_{N_g}^{(g)}\right)\right) = \sum_{g=1}^{G} \sum_{n=1}^{N_g} \sum_{i<i'} \ln L\left(\boldsymbol{\theta}; (y_{in}^{(g)}, y_{i'n}^{(g)})\right)$ (Muthén, 1989).

In the single-group setting, the data-generating model (Model 0) is tested against a model similar to Model 0 with the additional specification that the first three indicators of the second factor load on the first factor as well (Model 2). Variables are taken to have four response categories and two sample sizes, 200 and 500, are being considered. The empirical rates of Type I error are given in Figure 5 and in Table 5 (in the supplementary material). All three test statistics perform according to their asymptotic distribution at both significance levels except for $PLRT_{MV}$ which slightly under-rejects the null hypothesis at the 5% significance level when the sample size is 200. Furthermore, for both sample sizes, BIC_{PL} selects the right model with 100% success and performs better than AIC_{PL}. The percentage of the latter though is also relatively high; 88.9% and 89.7% for sample sizes 200 and 500, respectively.

In the multi-group setting, we follow the set-up of the example presented in Section 7 below. We generate two-group data from a five-factor model with 15 ordinal variables. Each factor is measured by a distinct set of three variables, the loadings of which are 0.6, 0.7, and 0.8 for all factors in both groups. Each indicator has four response categories and the thresholds are -1.25, 0, and 1.25 for all indicators in both groups. The mean vector and the covariance matrix of the factors, denoted respectively by $\boldsymbol{\alpha}^{(g)}$ and $\Psi^{(g)}$ in Section 2, differ in the two groups. Their values are given in Table 1. The covariance matrix of the measurement error terms is set equal across the groups, namely $\Theta^{(2)} = \Theta^{(1)} = diag\left(I - \Lambda\Psi^{(1)}\Lambda'\right)$. This means that the covariance matrix of the underlying variables $\Sigma^{(1)}$ in the first group is actually a correlation matrix while this is not the case for $\Sigma^{(2)}$ in the second group. The differences between $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are due solely to the differences between $\Psi^{(1)}$ and $\Psi^{(2)}$. The mean vector of the underlying variables is $\boldsymbol{\mu}^{(1)} = \Lambda\boldsymbol{\alpha}^{(1)} = \mathbf{0}$ for the first group and $\boldsymbol{\mu}^{(2)} = \Lambda\boldsymbol{\alpha}^{(2)}$ for the second group. The sample sizes of the groups are selected to be equal and three sizes are considered, 200, 500, and 1000 giving, in total, six experimental conditions. They are derived by crossing the three different sizes with two different tests of measurement invariance, loading-invariance and threshold-invariance given loading-invariance. Three models are

Figure 5: Empirical type I error rates for the test statistics, $\mathrm{PLRT_{MV}}$, $\mathrm{T_{DWLS-MV}}$, $\mathrm{T_{ULS-MV}}$, testing nested models (Model 2 vs Model 0) for data with 4 response categories, sample sizes 200, 500, and significance levels 5% and 1%; Model 2 allows three loadings to be estimated which are correctly fixed to 0 in Model 0; the bold horizontal lines represent the nominal significance level

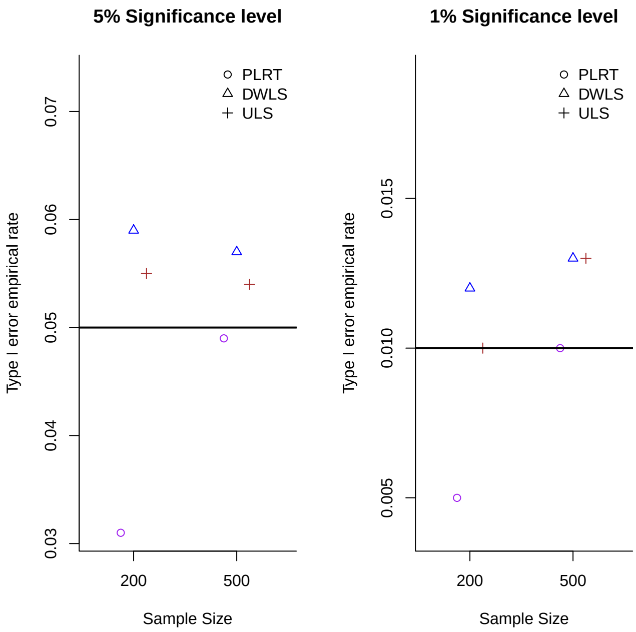


Table 1: The true values of the factor mean vectors and factor covariance matrices for the two-group generated data

Group 1	Group 2
$\boldsymbol{\alpha}^{(1)} = \mathbf{0}$	$\boldsymbol{\alpha}^{(2)} = \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}'$
$\Psi^{(1)} = \begin{pmatrix} 1 & & & & \\ 0.3 & 1 & & & \\ 0.3 & 0.4 & 1 & & \\ 0.3 & 0.4 & 0.5 & 1 & \\ 0.3 & 0.4 & 0.5 & 0.6 & 1 \end{pmatrix}$	$\Psi^{(2)} = \begin{pmatrix} 1.5 & & & & \\ 0.6 & 1.5 & & & \\ 0.6 & 0.8 & 1.5 & & \\ 0.6 & 0.8 & 0.9 & 1.5 & \\ 0.6 & 0.8 & 0.9 & 1.2 & 1.5 \end{pmatrix}$

fitted. Model A is the model with the minimum number of constraints needed for the model to be identified. As detailed in Millsap & Yun-Tein (2004), we have set: a) the mean and variance of the underlying variables equal to 0 and 1, respectively, in the first group; b) the loading and the first two thresholds of the first indicator of each latent variable equal between the groups; c) the first threshold of the rest of the indicators equal between the groups; and d) the factor means and variances of the first group equal to 0 and 1, respectively. Model B is the loading-invariant model which is actually Model A with cross-group equality constraints on all loadings (i.e. parameters in Λ matrix of Equation (1)). Model C is the loading and threshold invariant model which is Model B with cross-group equality constraints on all thresholds. For the loading-invariance test, Model B is compared to Model A, and for the threshold-invariance test given loading-invariance, Model C is compared to Model A.

Figure 6 and Table 6 (in the supplementary material) report the results for the test statistics. For the smallest group size, 200, $PLRT_{MV}$ under-rejects both hypotheses at both significance levels while the other two test statistics perform according to their asymptotic distribution. However, the performance of $PLRT_{MV}$ improves with the group size. In Table 2 we see that for all group sizes and model comparisons, AIC_{PL} selects the correct model with success close to 100% while BIC_{PL} always selects the right model.

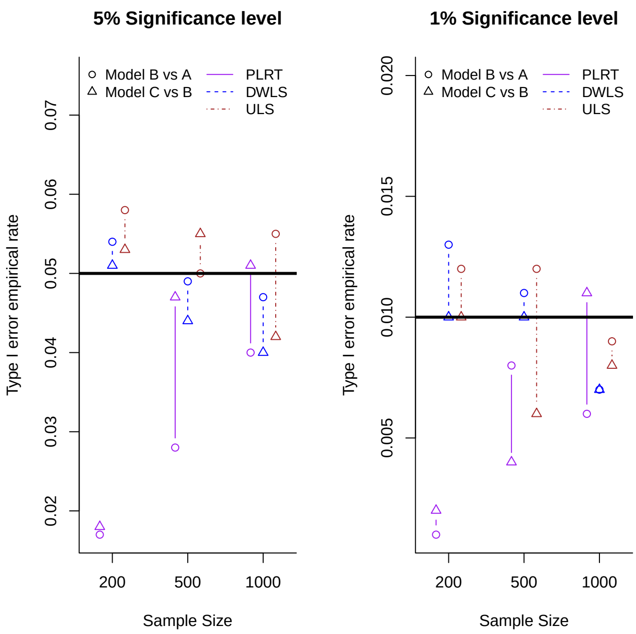
7.3 Conclusions based on the simulation results

The simulation results for both $PLRT_{MV}$ for nested models and $PLRT_{SEM-MV}$ for overall fit show acceptable levels of type I error and power. With respect to type I error, in most experimental conditions, the 95% confidence interval of the empirical rejection rates includes the nominal level (1% or 5%). If this is not the case, most probably in smaller sample sizes, the PLRT tests tend to under-reject a true null hypothesis which is preferable to over-rejection. However, their performance clearly improves with sample size. The power of $PLRT_{SEM-MV}$ depends on the

Table 2: Rates of AIC_{PL} and BIC_{PL} selecting the right model in two-group analysis for sample sizes 200, 500, 1000; Model A is the unconstrained model, Model B is the loading-invariant one, and Model C is the threshold- and loading-invariant model

N	200		500		1000	
	AIC_{PL}	BIC_{PL}	AIC_{PL}	BIC_{PL}	AIC_{PL}	BIC_{PL}
Model B vs A	96.6	100	96.6	100	97.0	100
Model C vs B	99.5	100	99.6	100	99.5	100
Model C vs A	99.8	100	99.6	100	99.6	100

Figure 6: Empirical type I error rates for the test statistics, $\mathrm{PLRT}_{\mathrm{MV}}$, $\mathrm{T}_{\mathrm{DWLS-MV}}$, $\mathrm{T}_{\mathrm{ULS-MV}}$, testing two-group nested models (Models A, B, and C) for variables with 4 response categories, sample sizes 200, 500, 1000, and significance levels 5% and 1%; Model A is the unconstrained model, Model B is the loading-invariant one, and Model C is the threshold- and loading-invariant model; the bold horizontal lines represent the nominal significance level; the vertical lines joining the symbols (circle, triangle) are used to distinguish among the three test statistics and do not represent a range of values



sample size and the size of misspecification; when either or both of them increase, the power improves substantially with a tendency to reach 1. With respect to both criteria, type I error and power, the performance of the PLRT tests is competitive to that of the tests derived under DWLS and ULS. The differences in performance of the three methods become negligible as the sample size increases. Finally, in our simulation results the model selection criteria AIC_{PL} and BIC_{PL} select the right model at least in 96% of the cases with BIC_{PL} always performing better than AIC_{PL}.

8 Application on trust in the police from the European Social Survey

We analyze fifteen questions from the UK and Ireland (sample sizes 2422 and 2576 respectively) from the European Social Survey (ESS), Round 5 (2010), section "Trust in the Police and Courts" (Section D in the questionnaire) (ESS Round 5, 2014; 2010). The data can be downloaded from the ESS webpage. The analysis consists of five latent variables measuring: "Trust in police effectiveness" (η_1), "Trust in police procedural fairness" (η_2), "Felt obligation to obey the police" (η_3), "Moral alignment with the police" (η_4) and, "Willingness to cooperate with the police" (η_5). Each latent variable is measured by three ordinal variables, the wording of which, along with the response categories, are given in Appendix A.4. The hypothesized model in each country is discussed in Jackson et al. (2012) and the relationships among the five constructs of interest are given below:

$$\begin{aligned}
\eta_3 &= \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3 \\
\eta_4 &= \beta_{41}\eta_1 + \beta_{42}\eta_2 + \zeta_4 \\
\eta_5 &= \beta_{51}\eta_1 + \beta_{53}\eta_3 + \beta_{54}\eta_4 + \zeta_5 \,.
\end{aligned}$$

The two-group SEM is fitted in `lavaan`. In principle, for valid cross-country comparisons, measurement invariance needs to hold. Three two-group models are fitted: Model A is the model with the minimum number of constraints needed to identify the two-group model (for details see Millsap & Yun-Tein, 2004); Model B is the loading-invariant model, which is Model A with cross-country equality constraints on all loadings; Model C is Model B with cross-country equality constraints on the thresholds of the first indicator of η_1 (namely, question D12). Model A is compared to Model B and Model B is compared to Model C. Table 3 presents the p-values of the three test statistics, $PLRT_{MV}$, T_{ULS-MV}, and $T_{DWLS-MV}$. All three test statistics fail to reject Model B (p-value>0.25). Model C is rejected at the 1% significance level by T_{ULS-MV} and $T_{DWLS-MV}$ (their p-values are less than 0.001) and at 5% by $PLRT_{MV}$ (p-value = 0.028). The table also reports

Table 3: Two-group analysis of ESS data: p-value of test statistics for nested models; AIC_{PL}, and BIC_{PL}

	Model B vs A	Model C vs B		Model A	Model B	Model C
$PLRT_{MV}$	0.48	0.028	AIC_{PL}	2209526	2209491	2209824
T_{ULS-MV}	0.31	0.000	BIC_{PL}	2226159	2225908	2225554
$T_{DWLS-MV}$	0.27	0.000				

Table 4: Two-group analysis of ESS data: values and p-values of overall-fit test statistics

	Model A	Model B	Model C
	value (p-value)	value (p-value)	value (p-value)
$PLRT_{MV}$	291.9 (0.000)	148.9 (0.000)	109.6 (0.000)
T_{ULS-MV}	735.0 (0.000)	670.8 (0.000)	740.0 (0.000)
$T_{DWLS-MV}$	1209.2 (0.000)	1255.4 (0.000)	1331.8 (0.000)

the AIC_{PL} and BIC_{PL} values of the three models. AIC_{PL} selects Model B while BIC_{PL} selects Model C. All test statistics including the PLRT for overall fit reject all three models (p-value<0.001). The values of test statistics for each model are given in Table 4.

All three test statistics reject all three fitted models. All three test statistics and model selection criteria for comparing models A and B suggest Model B. Interestingly, in the comparison of Model B with Model C, PLRT at 1% significance level agrees with BIC_{PL} in indicating Model C.

9 Conclusions

In this paper, we develop the pairwise likelihood ratio test (PLRT) for testing the overall fit of a structural equation model (SEM) for ordinal variables and for comparing nested models. Moreover, the composite likelihood versions of AIC and BIC are studied in SEM for ordinal variables. All the developed test statistics and model selection criteria are available in the R package `lavaan`. The type I error and power of the derived test statistics are investigated via a simulation study for models encountered in practice, such as a two-factor confirmatory model measured by 20 ordinal variables and a two-group model with 5 factors and 15 ordinal variables. The derived test statistics are also compared with test statistics derived under the three-stage diagonally weighted least squares (DWLS) and unweighted least squares (ULS). A real data example from the European Social Survey is used to illustrate their use in a two-group five-factor model fitted to the UK and Ireland data.

The PLRT for comparing nested models covers the case of nested models due to equality constraints among parameters and/or due to certain parameters being fixed equal to specific values. The PLRT for overall fit can be applied to models that do not only assume parametric structure on the polychoric correlations of the underlying variables but on the thresholds as well. Although the paper focuses on SEM for ordinal variables, the proposed methodology readily extends to SEM with mixed type variables (continuous and ordinal) and covariates.

The type I error and power of the PLRT statistics is quite satisfactory for the experimental conditions studied in this paper. The empirical type I error rates for PLRT is never higher than the nominal one. In most experimental conditions the 95% confidence interval (CI) of the empirical rate includes the nominal value of the significance level. It is mainly in the smaller sample size (200) that PLRT tends to under-reject a true null hypothesis. However, the performance improves with the sample size. The performance of the test statistics derived under DWLS and ULS with respect to type I error seems a bit better in the sense that in more experimental conditions the 95% CI of the empirical type I error rate includes the nominal significance level. However, whenever this is not the case, they tend to over-reject the null hypothesis. The performance of PLRT with respect to power improves substantially with the sample size and the misspecification size and is competitive to that of DWLS and ULS test statistics. The differences in their performances becomes negligible as the size of sample and/or misspecification increases. Furthermore, the model selection criteria, AIC_{PL} and BIC_{PL}, are found to select the right model with very high probability (at least 96% of the replications) with BIC_{PL} always performing better.

The paper considers the standard approach of mean-and-variance adjustment for the PLRT statistics. Further research should be conducted on studying other adjustments such as the one proposed by Pace et al. (2011). Moreover, the results regarding the overall fit PLRT statistic can be used in future research to derive fit indices that inspect the fit of the model on a subset of the observed variables. Such diagnostic tools are useful in practice since the overall fit test statistics often reject the hypothesized models.

10 Appendix

A.1. Proof for $PLRT\left(g\left(\boldsymbol{\theta}\right)\right)$

With $\hat{\boldsymbol{\theta}}$ being a PML estimator, it holds that $\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \to \mathcal{N}\left(\mathbf{0}, G^{-1}\left(\boldsymbol{\theta}_0\right)\right)$. Using the Delta method, $\sqrt{N}\left(g\left(\hat{\boldsymbol{\theta}}\right)-g\left(\boldsymbol{\theta}_0\right)\right) \to \mathcal{N}\left(\mathbf{0}, M\left(\boldsymbol{\theta}_0\right)G^{-1}\left(\boldsymbol{\theta}_0\right)\left[M\left(\boldsymbol{\theta}_0\right)\right]'\right)$, where $M\left(\boldsymbol{\theta}_0\right) = \frac{\partial}{\partial\boldsymbol{\theta}'}g\left(\boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. Under $H_0 : g\left(\boldsymbol{\theta}\right) = \mathbf{0}$, it holds $\sqrt{N}g\left(\hat{\boldsymbol{\theta}}\right) \to$

$\mathcal{N}\left(\mathbf{0}, M\left(\boldsymbol{\theta}_0\right) G^{-1}\left(\boldsymbol{\theta}_0\right)\left[M\left(\boldsymbol{\theta}_0\right)\right]'\right)$. Taking the second order Taylor expansion of $pl(\tilde{\boldsymbol{\theta}})$ around $\hat{\boldsymbol{\theta}}$ and since $\frac{\partial pl}{\partial \boldsymbol{\theta}'}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0}$ we get
$2\left(pl(\hat{\boldsymbol{\theta}}) - pl(\tilde{\boldsymbol{\theta}})\right) \simeq N(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'\left(-\frac{1}{N}\left.\frac{\partial^2 pl}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$.
Thus, $PLRT\left(g\left(\boldsymbol{\theta}\right)\right) \to N(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})' H(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$. Taking the first order Taylor expansion of $\frac{\partial pl}{\partial \boldsymbol{\theta}'}\big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$ around $\hat{\boldsymbol{\theta}}$ and since $\frac{\partial pl}{\partial \boldsymbol{\theta}'}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$ we get:

$$(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \to -\frac{1}{N} H^{-1}(\boldsymbol{\theta}_0) \left.\frac{\partial pl}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} . \tag{19}$$

Taking the first order Taylor expansion of $g\left(\tilde{\boldsymbol{\theta}}\right)$ around $\hat{\boldsymbol{\theta}}$ and since, under H_0, $g\left(\tilde{\boldsymbol{\theta}}\right) = 0$, it holds $g\left(\hat{\boldsymbol{\theta}}\right) \to -M\left(\hat{\boldsymbol{\theta}}\right)(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$. In the latter we substitute $\left(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\right)$ with (19) to get $g\left(\hat{\boldsymbol{\theta}}\right) \to \frac{1}{N} M\left(\hat{\boldsymbol{\theta}}\right) H^{-1}(\boldsymbol{\theta}_0) \frac{\partial pl}{\partial \boldsymbol{\theta}'}\big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$.
It holds $\frac{\partial pl}{\partial \boldsymbol{\theta}'}\big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}} = \left[M\left(\tilde{\boldsymbol{\theta}}\right)\right]' \boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is an $r \times 1$ vector of Lagrange multipliers.
Hence, $g\left(\hat{\boldsymbol{\theta}}\right) \to \frac{1}{N} M\left(\hat{\boldsymbol{\theta}}\right) H^{-1}(\boldsymbol{\theta}_0) \left[M\left(\tilde{\boldsymbol{\theta}}\right)\right]' \boldsymbol{\lambda}$ and
$\boldsymbol{\lambda} \to N\left\{M\left(\hat{\boldsymbol{\theta}}\right) H^{-1}(\boldsymbol{\theta}_0)\left[M\left(\tilde{\boldsymbol{\theta}}\right)\right]'\right\}^{-1} g\left(\hat{\boldsymbol{\theta}}\right)$.
In expression (19), we substitute $\frac{\partial pl}{\partial \boldsymbol{\theta}'}\big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$ and $\boldsymbol{\lambda}$ with the above results to get $(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \to -H^{-1}(\boldsymbol{\theta}_0)\left[M\left(\tilde{\boldsymbol{\theta}}\right)\right]'\left\{M\left(\hat{\boldsymbol{\theta}}\right) H^{-1}(\boldsymbol{\theta}_0)\left[M\left(\tilde{\boldsymbol{\theta}}\right)\right]'\right\}^{-1} g\left(\hat{\boldsymbol{\theta}}\right)$.
Under H_0, $(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) \to -H^{-1}(\boldsymbol{\theta}_0)\left[M\left(\boldsymbol{\theta}_0\right)\right]'\left[A\left(\boldsymbol{\theta}_0\right)\right]^{-1} g\left(\hat{\boldsymbol{\theta}}\right)$, where
$A\left(\boldsymbol{\theta}_0\right) = M\left(\boldsymbol{\theta}_0\right) H^{-1}(\boldsymbol{\theta}_0)\left[M\left(\boldsymbol{\theta}_0\right)\right]'$. Thus, $PLRT\left(g\left(\boldsymbol{\theta}\right)\right)$ can be written as follows
$PLRT\left(g\left(\boldsymbol{\theta}\right)\right) \to \left(\sqrt{N}[A\left(\boldsymbol{\theta}_0\right)]^{-1/2} g\left(\hat{\boldsymbol{\theta}}\right)\right)'\left(\sqrt{N}[A\left(\boldsymbol{\theta}_0\right)]^{-1/2} g\left(\hat{\boldsymbol{\theta}}\right)\right)$, where
$\sqrt{N}[A\left(\boldsymbol{\theta}_0\right)]^{-1/2} g\left(\hat{\boldsymbol{\theta}}\right) \to \mathcal{N}\left(\mathbf{0}, [A\left(\boldsymbol{\theta}_0\right)]^{-1/2} M\left(\boldsymbol{\theta}_0\right) G^{-1}\left(\boldsymbol{\theta}_0\right)\left[M\left(\boldsymbol{\theta}_0\right)\right]'\left[A\left(\boldsymbol{\theta}_0\right)\right]^{-1/2}\right)$.
Therefore, $PLRT\left(g\left(\boldsymbol{\theta}\right)\right) \to \sum_{i=1}^{r} \kappa_i u_i$, where u_i's are independent χ^2_1-distributed variables, and κ_i is the ith eigenvalue of matrix $[A\left(\boldsymbol{\theta}_0\right)]^{-1/2} M\left(\boldsymbol{\theta}_0\right) G^{-1}\left(\boldsymbol{\theta}_0\right)\left[M\left(\boldsymbol{\theta}_0\right)\right]'\left[A\left(\boldsymbol{\theta}_0\right)\right]^{-1/2}$.

A.2. Proof for $PLRT_{SEM}$

Before we consider the $PLRT_{SEM}$, we need to consider the PLRT statistics for two hypotheses of nested models. Firstly, consider the $PLRT(\boldsymbol{\varphi}_0)$ for the hypothesis $H_0 : \boldsymbol{\varphi} = \boldsymbol{\varphi}_0$ versus $H_1 : \boldsymbol{\varphi} \neq \boldsymbol{\varphi}_0$, where the SEM parameter $\boldsymbol{\theta}$ is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\omega}')'$, $\boldsymbol{\varphi}$ is the parameter vector of interest, $\boldsymbol{\omega}$ is the vector of nuisance parameters, and $\boldsymbol{\varphi}_0$ is a vector of real values. As we have already discussed in Section 4.1, this hypothesis is a special case of the hypothesis $H_0 : g\left(\boldsymbol{\theta}\right) = \mathbf{0}$, where $g\left(\boldsymbol{\theta}\right) = \boldsymbol{\varphi} - \boldsymbol{\varphi}_0$ and the matrices $A\left(\boldsymbol{\theta}_0\right)$ and $B\left(\boldsymbol{\theta}_0\right)$ are simplified to $H^{\varphi\varphi}\left(\boldsymbol{\theta}_0\right)$ and

$G^{\varphi\varphi}(\boldsymbol{\theta}_0)$, respectively. Using the result of the previous section, we conclude that

$$PLRT(\boldsymbol{\varphi}_0) \rightarrow \boldsymbol{v}'\boldsymbol{v} \tag{20}$$

where $\boldsymbol{v} = \sqrt{N}\left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1/2}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$. Since $\sqrt{N}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0) \rightarrow \mathcal{N}(\mathbf{0}, G^{\varphi\varphi}(\boldsymbol{\theta}_0))$, $\boldsymbol{v} \rightarrow \mathcal{N}\left(\mathbf{0}, \left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1/2} G^{\varphi\varphi}(\boldsymbol{\theta}_0)\left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1/2}\right)$.

Secondly, consider the $PLRT(\boldsymbol{\sigma}_0)$ for the hypothesis $H_0 : \boldsymbol{\sigma} = \boldsymbol{\sigma}_0$ versus $H_1 : \boldsymbol{\sigma} \neq \boldsymbol{\sigma}_0$, where $\boldsymbol{\vartheta}$ is the complete parameter vector of an unconstrained model, partitioned as $\boldsymbol{\vartheta} = (\boldsymbol{\sigma}', \boldsymbol{\tau}')'$, and $\boldsymbol{\sigma}_0$ is a vector of real values. Following the same reasoning as in $PLRT(\boldsymbol{\varphi}_0)$, it follows that:

$$PLRT(\boldsymbol{\sigma}_0) \rightarrow \boldsymbol{z}'\boldsymbol{z} \tag{21}$$

where $\boldsymbol{z} = \sqrt{N}\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1/2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0)$, and thus
$\boldsymbol{z} \rightarrow \mathcal{N}\left(\mathbf{0}, \left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1/2} G^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1/2}\right)$.

Now we return to $PLRT_{SEM}$. Let $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\varphi}_0', \tilde{\boldsymbol{\tau}}'_{\varphi_0})'$. Under H_0, it holds $\boldsymbol{\sigma}_0 = g(\boldsymbol{\varphi}_0)$ and thus $pl\begin{pmatrix} \boldsymbol{\sigma}_0 \\ \tilde{\boldsymbol{\tau}}_{\sigma_0} \end{pmatrix} = pl\begin{pmatrix} \boldsymbol{\varphi}_0 \\ \tilde{\boldsymbol{\tau}}_{\varphi_0} \end{pmatrix}$, i.e. $pl\left(\tilde{\boldsymbol{\vartheta}}\right) = pl\left(\tilde{\boldsymbol{\theta}}\right)$. This way, $PLRT_{SEM}$ can be written as
$PLRT_{SEM} = 2\left(pl\left(\hat{\boldsymbol{\vartheta}}\right) - pl\left(\hat{\boldsymbol{\theta}}\right)\right) = 2\left(pl\left(\hat{\boldsymbol{\vartheta}}\right) - pl\left(\tilde{\boldsymbol{\vartheta}}\right)\right) - 2\left(pl\left(\hat{\boldsymbol{\theta}}\right) - pl\left(\tilde{\boldsymbol{\theta}}\right)\right) = PLRT(\boldsymbol{\sigma}_0) - PLRT(\boldsymbol{\varphi}_0)$. Based on (20) and (21), $PLRT_{SEM} \rightarrow \boldsymbol{z}'\boldsymbol{z} - \boldsymbol{v}'\boldsymbol{v}$.

A.3. Proof for $Var\,(PLRT_{SEM})$

Since $PLRT_{SEM} \rightarrow \boldsymbol{z}'\boldsymbol{z} - \boldsymbol{v}'\boldsymbol{v}$ where $\boldsymbol{z} = \sqrt{N}\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1/2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0)$ and $\boldsymbol{v} = \sqrt{N}\left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1/2}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$, it follows:

$$Var\,(PLRT_{SEM}) \rightarrow Var\,(\boldsymbol{z}'\boldsymbol{z}) + Var\,(\boldsymbol{v}'\boldsymbol{v}) - 2Cov\,(\boldsymbol{z}'\boldsymbol{z}, \boldsymbol{v}'\boldsymbol{v})$$

with $Var\,(\boldsymbol{z}'\boldsymbol{z}) = 2\mathrm{tr}\left(G^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1} G^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1}\right)$, $Var\,(\boldsymbol{v}'\boldsymbol{v}) = 2\mathrm{tr}\left(G^{\varphi\varphi}(\boldsymbol{\theta}_0)\left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1} G^{\varphi\varphi}(\boldsymbol{\theta}_0)\left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1}\right)$, and the calculations for $Cov\,(\boldsymbol{z}'\boldsymbol{z}, \boldsymbol{v}'\boldsymbol{v})$ are shown below. Under H_0, $\boldsymbol{\sigma}_0 = g(\boldsymbol{\varphi}_0)$, so it can be written as
$\boldsymbol{z}'\boldsymbol{z} = N(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0)'\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_0) =$
$= N(g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0))'\left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1}(g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0))$. Therefore,
$Cov\,(\boldsymbol{z}'\boldsymbol{z}, \boldsymbol{v}'\boldsymbol{v}) = Cov\left[N(g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0))' A\,(g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0)), N(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)' B\,(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)\right]$,
where $A = \left[H^{\sigma\sigma}(\boldsymbol{\vartheta}_0)\right]^{-1}$ and $B = \left[H^{\varphi\varphi}(\boldsymbol{\theta}_0)\right]^{-1}$, both being symmetric matrices. Based on the first-order Taylor expansion of $g(\hat{\boldsymbol{\varphi}})$ around $g(\boldsymbol{\varphi}_0)$:
$g(\hat{\boldsymbol{\varphi}}) \simeq g(\boldsymbol{\varphi}_0) + \frac{\partial}{\partial \boldsymbol{\varphi}} g(\boldsymbol{\varphi})\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_0}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$, where $\frac{\partial}{\partial \boldsymbol{\varphi}} g(\boldsymbol{\varphi})\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_0}$. Let $C = \frac{\partial}{\partial \boldsymbol{\varphi}} g(\boldsymbol{\varphi})\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_0}$.

Thus, $g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0) \simeq C(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$ and $(g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0))' A (g(\hat{\boldsymbol{\varphi}}) - g(\boldsymbol{\varphi}_0)) \simeq (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)' D (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)$, where $D = C'AC$ and is symmetric because A is symmetric. The covariance expression can now be written as:

$$\begin{aligned} Cov(\boldsymbol{z}'\boldsymbol{z}, \boldsymbol{v}'\boldsymbol{v}) &\simeq Cov\left[N(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)' D (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0), N(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)' B (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0)\right] \\ &= 2\mathrm{tr}\left(DG^{\varphi\varphi}BG^{\varphi\varphi}\right) \\ &= 2\mathrm{tr}\left(\left(\frac{\partial}{\partial\boldsymbol{\varphi}} g(\boldsymbol{\varphi})\bigg|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_0}\right)' [H^{\sigma\sigma}]^{-1} \frac{\partial}{\partial\boldsymbol{\varphi}} g(\boldsymbol{\varphi})\bigg|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_0} G^{\varphi\varphi} [H^{\varphi\varphi}]^{-1} G^{\varphi\varphi}\right). \end{aligned}$$

The expression of the first line is equal to that of the second line by using the result, proved in Magnus (1978), that if $\mathbf{t} \to N(\mathbf{0}, V)$, then $Cov(\mathbf{t}'D\mathbf{t}, \mathbf{t}'B\mathbf{t}) = 2\mathrm{tr}(DVBV)$. (This result can also be used for the computations of $Var(\boldsymbol{z}'\boldsymbol{z})$ and $Var(\boldsymbol{v}'\boldsymbol{v})$.) The expressions of the last two lines above are equal by simply substituting the matrices D and B with their equivalence.

A.4. Questions on trust in the police, European Social Survey, Round 5.

Trust in police effectiveness
D12. Based on what you have heard or your own experience how successful do you think the police are at preventing crimes in [country] where violence is used or threatened?
D13. How successful do you think the police are at catching people who commit house burglaries in [country]?
D14. If a violent crime were to occur near to where you live and the police were called, how slowly or quickly do you think they would arrive at the scene?

Trust in police procedural fairness
D15. Based on what you have heard or your own experience how often would you say the police generally treat people in [country] with respect?
D16. About how often would you say that the police make fair, impartial decisions in the cases they deal with?
D17. When dealing with people in [country], how often would you say the police generally explain their decisions and actions when asked to do so?

Felt obligation to obey the police
To what extent is it your duty to…
D18. …back the decisions made by the police even when you disagree with them?
D19. …do what the police tell you even if you don't understand or agree with the reasons?

D20. … do what the police tell you to do, even if you don't like how they treat you?

Moral alignment with the police
D21. The police generally have the same sense of right and wrong as I do.
D22. The police stand up for values that are important to people like me.
D23. I generally support how the police usually act.

Willingness to cooperate with the police
D40. Imagine that you were out and saw someone push a man to the ground and steal his wallet. How likely would you be to call the police?
D41. How willing would you be to identify the person who had done it?
D42. And how willing would you be to give evidence in court against the accused?

Response Scales
11-point for questions D12-D14, D18-D20; 0 denotes "Extremely Unsuccessful"/"Extremely slowly"/"Not at all my duty"; 10 denotes "Extremely Successful"/"Extremely quickly"/ "Completely my duty".
4-point for questions D15-D17 and D40-D42. For D15-D17, 1 denotes "Not at all often" and 4 "Very often". For D40-D42, 1 denotes "Not at all likely" and 4 "Very likely".
5-point for questions D21-D23, 1 denotes "Agree strongly" and 5 "Disagree strongly".
The extra response category: "Violent crimes never occur near to where I live" in D14 is treated as missing in our analysis.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. Wiley, 2nd ed.

Ansari, A., & Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika*, *65*(4), 475–496.

Ansari, A., & Jedidi, K. (2002). Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika*, *67*(1), 49–78.

Arminger, G., & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In I. R. Langeheine, & J. Rost (Eds.) *Latent Trait and Latent Class Models*. New York: Plenum.

Asparouhov, T., & Muthén, B. (2006). Robust chi-square difference testing with mean and variance adjusted test statistics. *Mplus Web Notes: No. 10*.
URL http://www.statmodel.com/download/webnotes/webnote10.pdf

Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction.
URL https://www.statmodel.com/download/WLSMV_new_chi21.pdf

Bartholomew, D., Knott, M., & Moustaki, I. (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley series in Probability and Statistics, 3rd ed.

Bellio, R., & Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling*, *5*, 217–227.

Bentler, P. M. (2006). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society Series B*, *36*, 192–236.

Bhat, C. R., Varin, C., & Ferdous, N. (2010). *Maximum Simulated Likelihood Methods and Applications (Advances in Econometrics, Volume 26)*, chap. A Comparison of the Maximum Simulated Likelihood and Composite Marginal Likelihood Estimation Approaches in the Context of the Multivariate Ordered Response Model, (pp. 65–106). Emerald Group Publishing Limited.

Bollen, K., & Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. Wiley Series in Probability and Mathematical Statistics. New York.

De Leon, A. R. (2005). Pairwise likelihood approach to grouped continuous model and its extension. *Statistics & Probability Letters*, *75*, 49–57.

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, *65*(3), 457–487.

ESS (2010). ESS Round 5: European Social Survey Round 5 Data. Data file edition 3.2. *Norwegian Social Science Data Services, Norway, Data Archive and distributor of ESS data*.

ESS (2014). Round 5: European Social Survey: ESS-5 Documentation Report. Edition 3.2. *Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services*.

Fan, W., & Hancock, G. R. (2012). Robust means modeling: An alternative for hypothesis testing of independent means under variance heterogeneity and nonnormality. *Journal of Educational and Behavioral Statistics*, *37*, 137–156.

Feddag, M.-L., & Bacci, S. (2009). Pairwise likelihood for the longitudinal mixed Rasch model. *Computational Statistics and Data Analysis*, *53*, 1027–1037.

Fieuws, S., & Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, *62*, 424–431.

Gao, X., & Song, P. X. (2010). Composite likelihood Bayesian information criteria for model selection in high dimensional data. *Journal of the American Statistical Association*, *105*(492), 1531–1540.

Heagerty, P. J., & Lele, S. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, *93*, 1099–1111.

Jackson, J., Hough, M., Bradford, B., Hohl, K., & Kuha, J. (2012). Policing by consent: Topline results (UK) from Round 5 of the European social survey. *ESS Country Specific Topline Results Series 1*.

Joe, H., & Lee, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *Journal of Multivariate Analysis*, *100*, 670–685.

Jöreskog, K., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. Marcoulides, & R. Schumacker (Eds.) *Advanced Structural Equation Modeling: Issues and Techniques*, (pp. 57–88). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183–202.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.

Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, *24*, 387–404.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, *59*, 381–389.

Jöreskog, K. G. (2002). Structural equation modeling with ordinal variables using LISREL.
URL `http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf`

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347–387.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 User's Reference Guide*. Chicago, IL: Scientific Software International.

Katsikatsou, M. (2013). *Composite Likelihood Estimation for Latent Variable Models with Ordinal and Continuous or Ranking Variables*. Ph.D. thesis, Uppsala University, Sweden.

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics and Data Analysis*, *56*, 4243–4258.

Kenward, M. G., & Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, *13*(3), 236–247.

Lee, S.-Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. Wiley Series in Probability and Statistics.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. (1990a). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters*, *9*, 91–97.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1990b). A three-stage estimation procedure for structural equation models with polytomous variables. *Psychometrika*, *55*, 45–51.

Lee, S.-Y., Poon, W.-Y., & Bentler, P. M. (1992). Structural equation models with continuous and polytomous variables. *Psychometrika*, *57*, 89–105.

Lele, S. R. (2006). Sampling variability and estimates of density dependence: A composite likelihood approach. *Ecology*, *87*, 189–202.

Lele, S. R., & Taper, M. L. (2002). A composite likelihood approach to (co)variance components estimation. *Journal of Statistical Planning and Inference*, *103*, 117–135.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.

Liu, J. (2007). *Multivariate Ordinal Data Analysis with Pairwise Likelihood and its Extension to SEM*. Ph.D. thesis, University of California, Los Angeles.
URL `http://statistics.ucla.edu/theses/uclastat-dissertation-2007:7`

Magnus, J. (1978). The moments of products of quadratic forms in normal variables. Tech. Rep. Technical Report AE4/78, Institute of Actuarial Science and Econometrics, Amsterdam University.
URL `http://www.janmagnus.nl/papers/JRM003.pdf`

Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2^n contingency tables: A unified approach. *Journal of the American Statistical Association*, *100*, 1009–1020.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*(4), 713–732.

Millsap, E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variables indicators. *Psychometrika*, *49*, 115–132.

Muthén, B. (1989). Multi-group structural modelling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, *42*, 55–62.

Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. In K. Bollen, & J. Long (Eds.) *Testing structural equation models*, (pp. 205–234). Sage Publications, Newbury Park.

Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes 4*.
URL `http://www.statmodel.com/download/webnotes/CatMGLong.pdf`

Muthén, B., du Toit, S., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.
URL http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (2010). *Mplus 6*. Muthén and Muthén, Los Angeles.

Pace, L., Salvan, A., & Sartori, N. (2011). Adjusting composite likelihood ratio statistics. *Statistica Sinica*, *21*, 129–148.

Palomo, J., Dunson, D. B., & Bollen, K. (2007). *Handbook of Computing and Statistics with Applications Vol. 1: Handbook of Latent Variable and Related Models*, chap. Chapter 8, Bayesian Structural Equation Modeling, (pp. 163–188). Elsevier.

Poon, W.-Y., & Lee, S.-Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika*, *52*, 409–430.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.r-project.org

Raftery, A. (1993). Bayesian model selection in structural equation models. In K. Bollen, & J. Long (Eds.) *Testing Structural Equation Models*. Sage, Newbury Park, CA.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48 (2)*, 1–36.
URL http://www.jstatsoft.org/v48/i02/paper

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., & Merkle, E. (2012). *Package lavaan*.
URL http://cran.r-project.org/web/packages/lavaan/lavaan.pdf

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.) *Innovations in Multivariate Statistical Analysis. A Festschrift for Heinz Neudecker*, (pp. 233–247). London: Kluwer Academic Publishers.

Satorra, A., & Bentler, P. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, (pp. 308–313).

Satorra, A., & Bentler, P. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. Clogg (Eds.) *Latent Variable Analysis: Applications to Developmental Research*, (pp. 399–419). Sage Publications, Thousand Oaks, CA.

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514.

Satorra, A., & Bentler, P. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248.

Savalei, V., & Kolenikov, S. (2008). Constrained vs. unconstrained estimation in structural equation modeling. *Psychological Methods*, *13*, 150–170.

Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Chap.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, *92*, 1–28.

Varin, C., Høst, G., & Øivind, S. (2005). Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics and Data Analysis*, *49*, 1173–1191.

Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, *21*, 1–41.

Varin, C., & Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, *92*, 519–528.

Varin, C., & Vidoni, P. (2006). Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics and Data Analysis*, *51*, 2365–2373.

Vasdekis, V., Cagnone, S., & Moustaki, I. (2012). A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika*, *77*, 425–441.

Wall, M., & Amemiya, Y. (2000). Estimation of polynomial structural equation models. *Journal of the American Statistical Association*, *95*, 929–940.

Xi, N. (2011). *A Composite Likelihood Approach for Factor Analyzing Ordinal Data*. Ph.D. thesis, The Ohio State University.

Zhao, Y., & Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, *33*, 335–356.